<u>**AMENDMENTS TO THE CLAIMS**</u>

This listing of claims will replace all prior versions of claims in the application:

Listing of Claims:

1.   (Currently Amended) A <u>computer-implemented</u> spam detection system comprising:

a message parsing component that identifies features relating to at least a portion of origination information of a message; and

a feature pairing component that combines the features into useful pairs, the features of the pairs are evaluated for consistency with respect to one another to determine if the message is spam.

2.   (Previously Presented) The system of claim 1, each pair comprises at least one of the following:

at least one of a domain name and a host name in a MAIL FROM command;

at least one of a domain name and a host name in a HELO COMMAND;

at least one of an IP address and a subnet in a Received from header;

at least one of a domain name and a host name in a Display name;

at least one of a domain name and a host name in a Message From line; and

at least one time zone in a last Received from header.

3.   (Previously Presented) The system of claim 2, the domain name is derived from the host name.

4.   (Previously Presented) The system of claim 2, the subnet comprises one or more IP addresses that share a first number of bits in common.

5.   (Previously Presented) The system of claim 1, a useful pair is any one of a domain name and a host name from a Message From and from a HELO command.

6.   (Previously Presented) The system of claim 1, a useful pair is a Display name domain name and host name and a Message From domain name and host name.

7.   (Previously Presented) The system of claim 1, a useful pair is any one of a domain name and a host name in a Message From and any one of a Received from IP address and subnet.

8.   (Previously Presented) The system of claim 1, a useful pair is a sender's alleged time zone and a Message From domain name.

9.   (Previously Presented) The system of claim 1, a useful pair comprises a sender's type of mailing software and any one of a domain name, host name and user name derived from one of an SMTP command and a message header.

10.   (Previously Presented) The system of claim 1, origination information comprises SMTP commands, the SMTP commands comprise a HELO command, a MAIL FROM command, and a DATA command.

11.   (Previously Presented) The system of claim 10, the DATA command comprises a Message From line, sender's alleged time zone, and sender's mailing software.

12.   (Original) The system of claim 1, further comprising a component that applies one or more heuristics consistently to mail messages to obtain consistent feature pairing.

13.   (Withdrawn) A spam detection system comprising:
        a character sequencing component that analyzes a portion of a message via searching for particular character sequences that are indicative of spam, the particular sequences are not restricted to whole words; and
        a feature generating component that generates features relating to the character sequences of any length, the features are analyzed to detect at least one of intentional character substitutions, insertions, or misspellings indicative of spam.

14.   (Withdrawn) The system of claim 0, the feature generating component generates features for each run of characters up to a maximum character run length.

15.   (Withdrawn) The system of claim 0, the feature generating component generates features for substantially all character sequences up to some length n.

16.   (Withdrawn) The system of claim 0, the character sequences comprise at least one of letters, numbers, punctuation, symbols, and characters of foreign languages.

17.   (Withdrawn) The system of claim 0, the particular character sequences comprise at least one of random letters, symbols, and punctuation as chaff at any one of a beginning and end of at least one of a subject line of a message and a message body.

18.   (Withdrawn) The system of claim 17, random character sequences comprise character n-grams which are indicative of spam-like messages.

19.   (Withdrawn) The system of claim 18, the character n-grams are located in at least one of From address, subject line, text body, html body, and attachments.

20.   (Withdrawn) The system of claim 18, the character n-grams are position dependent.

21.   (Withdrawn) The system of claim 0, the portion of the message comprising at least one of foreign language text, Unicode character types, and other character types not common to English

22.   (Withdrawn) The system of claim 21, the foreign language text comprises substantially non-space separated words.

23.   (Withdrawn) The system of claim 22, n-grams are used for characters not typically separated by spaces.

24.   (Withdrawn) The system of claim 0, further comprising a component that extracts character sequences obfuscated by punctuation using a pattern-match technique.

25.   (Withdrawn) A spam detection system comprising:

a character sequencing component that analyzes a portion of a message via searching for instances of a string of random characters that are indicative of the message being spam; and

a feature generating component that generates features corresponding to the instances of random character strings to facilitate determining an entropy measurement for each string, the entropy measurement is used to indicate the message as being spam or not spam.

26.   (Cancelled).

27.   (Withdrawn) The system of claim 25, the system measures a value correlated with entropy.

28.   (Withdrawn) The system of claim 27, a high value correlated with entropy is indicative of spam.

29.   (Withdrawn) The system of claim 28, the value correlated with entropy is the actual entropy $-\log_2 P(abc...z)$

30.   (Withdrawn) The system of claim 27, the average entropy of a character string is used.

31.   (Withdrawn) The system of claim 25, the string of random characters is chaff.

32.   (Withdrawn) The system of claim 27, the relative entropy compares the entropy measurement at any one of a beginning and end of at least one of a subject line and message body with the entropy measurement at a middle of at least one of the subject line and message body.

33.   (Withdrawn) A spam detection system comprising:

a component that analyzes substantially all features of a message header in connection with training a machine learning spam filter, the component generates feature pairs; and

a spam filter that detects spam based at least in part on a comparison of the feature pairs.

34.   (Withdrawn) The system of claim 33, the features of the message header comprise at least one of a presence and absence of at least one message header type, the message header types comprising X-Priority, mail software, and headers line for unsubscribing.

35.   (Withdrawn) The system of claim 34, the features of the message header further comprise content associated with at least one message header type.

36.   (Withdrawn) The system of claim 33, further comprising:

a component that analyzes at least a portion of a message for images and related image information; and

a component that generates features relating to any one of the images and related image information.

37.   (Withdrawn) The system of claim 36, the image information comprises image size, image quantity, location of image, image dimensions, and image type.

38.   (Withdrawn) The system of claim 36, the image information comprises the presence of a first URL and a second URL such that the image is inside of a hyperlink.

39.   (Withdrawn) The system of claim 38, the message comprises a tag pattern having the form of <A HREF="the first URL"><IMG SRC="the second URL"></A>.

40.   (Withdrawn) The system of claim 36, the features are used in connection with training a machine learning filter.

41.   (Withdrawn) The system of claim 33, further comprising a component that analyzes a message for HTML attributes and location of HTML attributes as they appear in a tag pattern.

42.   (Currently Amended) A computer-implemented method that facilitates generating features for use in spam detection comprising:

     receiving at least one message;

     parsing at least a portion of a message to generate one or more features;

     combining at least two features into pairs, each pair of features creates at least one additional feature, the features of each pair coinciding with one another;

     using the pairs of features to train a machine learning spam filter regarding acceptable or unacceptable pairs; and

     detecting a spam e-mail based at least in part on comparing one or more pairs of features in the e-mail to at least one pair in the machine learning spam filter.

43.   (Previously Presented) The method of claim 42, the at least a portion of the message being parsed corresponds to origination information of the message.

44.   (Previously Presented) The method of claim 42, each pair comprises at least one of the following:

     at least one of a domain name and a host name in a MAIL FROM command;

     at least one of a domain name and a host name in a HELO COMMAND;

     at least one of an IP address and a subnet in a Received from header;

     at least one of a domain name and a host name in a Display name;

     at least one of a domain name and a host name in a Message From line; and

     at least one time zone in a last Received from header.

45.   (Previously Presented) The method of claim 0, the domain name is derived from the host name.

46.   (Previously Presented) The method of claim 42, the pair of features is a Display name domain name and host name and a Message From domain name and host name.

47.   (Previously Presented) The method of claim 42, a useful pair is any one of a domain name and a host name from a Message From and from a HELO command.

48.   (Previously Presented) The method of claim 42, the pair of features is any one of a domain name and a host name in a Message From and any one of a Received from IP address and subnet.

49.   (Previously Presented) The method of claim 42, the pair of features is a sender's alleged time zone and a Message From domain name.

50.   (Previously Presented) The method of claim 42, the pair of features comprises a sender's type of mailing software and any one of a domain name, host name and display name derived from one of an SMTP command and a message header.

51.   (Original) The method of claim 42, further comprising selecting one or more most useful pairs of features to train the machine learning filter.

52.   (Previously Presented) The method of claim 42, the detecting a spam e-mail based at least in part on one of:

　　　　receiving new messages;

　　　　generating pairs of features based on origination information in the messages;

　　　　passing the pairs of features through the machine learning filter; and

　　　　obtaining a verdict as to whether at least one pair of features indicates that the message is more likely to be spam.

53.   (Withdrawn) A method that facilitates generating features for use in spam detection comprising:

        receiving one or more messages;

        walking through at least a portion of the message to create features for each run of characters of any run length; and

        training a machine learning filter on spam-indicative features using at least a portion of the created features, the filter subsequently identifies at least one spam-indicative feature in a message regardless of whitespace or extraneous characters in the features of the message.

54.   (Withdrawn) The method of claim 53, further comprising generating features relating to a position of at least one run of characters.

55.   (Withdrawn) The method of claim 54, the position comprises any one of a beginning of a message body, an end of a message body, a middle of a message body, a beginning of a subject line, an end of a subject line, and a middle of a subject line.

56.   (Withdrawn) The method of claim 53, the features are created for a run of characters up to length n.

57.   (Withdrawn) The method of claim 53, the features are created for sub-lengths of runs of characters.

58.   (Withdrawn) The method of claim 53, the run of characters comprise character n-grams.

59.   (Withdrawn) The method of claim 53, further comprising calculating an entropy of one or more run of characters and employing the calculated entropy as a feature in connection with training a spam filter.

60.   (Withdrawn) The method of claim 59, the entropy is at least one of high entropy, average entropy, and relative entropy.

61.   (Withdrawn) The method of claim 60, the average entropy is the entropy per character of a particular run of characters.

62.   (Withdrawn) The method of claim 60, the relative entropy is a comparison of the entropy of a particular run of characters at a first location relative to the entropy of a particular run of characters at a second location of the message.

63.   (Withdrawn) The method of claim 62, the first and second locations comprise a beginning of a subject line, a middle of a subject line, and an end of a subject line, the first location is not the same as the second location when determining the relative entropy for any given run of characters.

64.   (Withdrawn) The method of claim 62, the first and second locations comprise a beginning of a message, a middle of a message, and an end of a message, the first location is not the same as the second location when determining the relative entropy for any given run of characters.

65.   (Withdrawn) The method of claim 53, further comprising employing the machine learning filter after it is trained to detect spam by performing the following:

   receiving new messages;

   generating features based at least one of runs of characters and entropy determinations of runs of characters in the messages;

   passing the features through the machine learning filter; and

   obtaining a verdict as to whether the features indicate that the message is more likely to be spam.

66.   (Withdrawn) A method that facilitates generating features for use in spam detection comprising:

      receiving one or more messages;

      analyzing substantially all features of a message header, the features are compared to determine inconsistencies indicative of spam; and

      training a machine learning filter using the analyzed features.

67.   (Withdrawn) The method of claim 66, further comprising analyzing substantially all features based on image information in the message.

68.   (Withdrawn) A computer readable medium comprising the method of claim 42.

69.   (Withdrawn) A computer readable medium comprising the method of claim 53.

70.   (Withdrawn) A computer-readable medium having stored thereon the following computer executable components:

      a component that identifies features relating to at least a portion of origination information of a message; and

      a component that combines the features into useful pairs, the pairs are evaluated for consistency with respect to one another to determine if the message is spam.

71.   (Withdrawn) The computer readable medium of claim 70, further comprising:

      a component that analyzes a portion of a message via searching for particular character sequences that are indicative of spam, the particular sequences are not restricted to whole words; and

      a component that generates features relating to the character sequences of any length.

72.   (Withdrawn) The computer readable medium of claim 70, further comprising:

      a component that analyzes a portion of a message via searching for instances of a string of random characters that are indicative of the message being spam.

73.   (Currently Amended) A <u>computer-implemented</u> system that facilitates generating features for use in spam detection comprising:

      means for receiving at least one message;

      means for parsing at least a portion of a message to generate one or more features;

      means for combining at least two features into pairs, the pairs are evaluated against each other for consistency; and

      means for using the pairs of features to train a machine learning spam filter.

74.   (Withdrawn) A system that facilitates generating features for use in spam detection comprising:

      means for receiving one or more messages;

      means for walking through at least a portion of the message to create features for each run of characters of any run length; and

      means for training a machine learning filter on spam indicative features using at least a portion of the created features.

75.   (Withdrawn) The system of claim 74, further comprising means for calculating an entropy of one or more run of characters and employing the calculated entropy as a feature in connection with training a spam filter.

<u>7</u>